Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Infection, Genetics and Evolution 110 (2023) 105426

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Genetic susceptibility to severe COVID-19

Claudio Cappadona [a], Valeria Rimoldi [a,b], Elvezia Maria Paraboschi [a,b], Rosanna Asselta [a,b,*]

[a] Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini 4, Pieve Emanuele - Milan 20072, Italy
[b] IRCCS Humanitas Research Hospital, Via Manzoni 56, Rozzano, Milan 20089, Italy

ARTICLE INFO

Keywords:
SARS-CoV-2
COVID-19
Severity
Genetics
Genome-wide association study
Mutation

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiologic agent of the coronavirus disease 2019 (COVID-19) pandemic. Clinical manifestations of the disease range from an asymptomatic condition to life-threatening events and death, with more severe courses being associated with age, male sex, and comorbidities. Besides these risk factors, intrinsic characteristics of the virus as well as genetic factors of the host are expected to account for COVID-19 clinical heterogeneity.

Genetic studies have long been recognized as fundamental to identify biological mechanisms underlying congenital diseases, to pinpoint genes/proteins responsible for the susceptibility to different inherited conditions, to highlight targets of therapeutic relevance, to suggest drug repurposing, and even to clarify causal relationships that make modifiable some environmental risk factors. Though these studies usually take long time to be concluded and, above all, to translate their discoveries to patients' bedside, the scientific community moved really fast to deliver genetic signals underlying different COVID-19 phenotypes.

In this Review, besides a concise description of COVID-19 symptomatology and of SARS-CoV-2 mechanism of infection, we aimed to recapitulate the current literature in terms of host genetic factors that specifically associate with an increased severity of the disease.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiologic agent of the coronavirus disease 2019 (COVID-19) pandemic (Zhu et al., 2020). The first cases of this novel disease were described as suffering from severe pneumonia with unknown cause, and were reported in Wuhan (China) in late 2019. Soon after, while it was rapidly spreading to the rest of the worldwide population, SARS-CoV-2 was isolated and its 30-kb RNA genome sequenced (Wu et al., 2020); the World Health Organization (WHO) finally declared the pandemic on March 11, 2020. Three years after the description of the first case, and notwithstanding the introduction of vaccines, countries have reported a total of 660,317,813 confirmed infections, including 6,689,978 deaths [https://coronavirus.jhu.edu/ – accessed on December 31st, 2022].

It is now well established that COVID-19 is a systemic disease, with a broad clinical spectrum varying from the lack of symptoms (approximately one-third of infected are asymptomatic), to critical disease that can exit in death (Fig. 1A) (Driggin et al., 2020; Mehta et al., 2020; Terpos et al., 2020). In particular, the first clinical presentations of the infection are similar to symptoms caused by other respiratory viruses such as influenza/parainfluenza viruses, and include fever, cough, and fatigue. Less common signs comprise headache, sore throat, myalgia, arthralgia, diarrhea, vomiting, and changes in smell (anosmia, hyposmia) and taste (ageusia, dysgeusia) (da Silva et al., 2022). In a large study, counting >70 thousands subjects with COVID-19, the vast majority of cases (81%) presented the above-mentioned mild/moderate

Abbreviations: ACE2, Angiotensin I converting enzyme 2; CCR2, C-C chemokine receptor type 2; CCR9, CC motif chemokine receptor 9; CXCR6, C-X motif chemokine receptor 6; CI, Confidence interval; COVID-19, Coronavirus disease 2019; EMT, Epithelial-mesenchymal transition response; GHS, Geisinger Health System; GRS, Genetic risk score; GWAS, Genome-wide association study; HGI, Host Genetics Initiative; ICU, Intensive care unit; IFIH1, Interferon induced with helicase C domain 1; IFN, Interferon; IRF, Interferon regulatory factor; ISG, Interferon-stimulated gene; LD, Linkage disequilibrium; LZTFL1, Leucine zipper transcription factor like 1; NGS, Next-generation sequencing; OR, Odds ratio; PMBB, Penn Medicine BioBank; PRRs, Pattern recognition receptors; PRS, Polygenic risk score; RIG-I, Retinoic acid-inducible gene I; SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2; SLC6A20, Sodium imino-acid transporter; SNP, single-nucleotide polymorphism; TLR, Toll-like receptor; TMPRSS2, Transmembrane protease, serine 2; TLR7, Sensor toll-like receptor; UKB, UK Biobank; WGS, Whole-genome sequencing; WHO, World Health Organization.

* Corresponding author at: Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini 4, Pieve Emanuele 20090, Italy.
E-mail address: rosanna.asselta@hunimed.eu (R. Asselta).

https://doi.org/10.1016/j.meegid.2023.105426
Received 3 February 2023; Received in revised form 14 March 2023; Accepted 16 March 2023
Available online 17 March 2023

symptoms, a fraction (14%) progressed towards severe pneumonia that required respiratory support, whereas 5% of patients suffered from critical manifestations, including respiratory failure, septic shock, and multiple organ dysfunction or failure (Chen et al., 2020; Wu and McGoogan, 2020), which also represent the most common causes of COVID-19 death. To date, fatality ratios (defined as the number of deaths divided by the number of confirmed cases) for the 20 countries most affected by COVID-19 worldwide range from 0.1% registered in South Korea to 4.9% reported in Peru (Fig. 1B) [https://coronavirus.jhu.edu/data/mortality – accessed on December 28th, 2022]. Major risk factors for the development of COVID-19 severe complications (and death) comprise age above 65 years, male sex, as well as comorbidities, such as hypertension, diabetes, chronic pulmonary, kidney, or liver disease, immunodeficiencies, cancer, cardiovascular disease, and obesity (Grasselli et al., 2020; Petrilli et al., 2020). Finally, it is now clear that COVID-19 patients can progress into a post-COVID-19 syndrome (also known as long COVID-19 or long-haul COVID-19) that is characterized by symptoms and abnormalities persisting beyond 3 months of COVID-19 onset (Moreno-Pérez et al., 2021).

As for SARS-CoV-2 pathophysiologic mechanisms, virus attachment and entry into epithelial nasopharynx and subsequently into alveolar cells are mediated by the presence of the angiotensin I converting enzyme 2 (ACE2), a strategy that is common with SARS-CoV (the first-appearing severe acute respiratory syndrome coronavirus) (Tsang et al., 2003). ACE2 is highly/moderately expressed in many human tissues both at the mRNA and at the protein levels [The Genotype-Tissue Expression (GTEx) project, https://gtexportal.org/home/; The Human Protein Atlas, https://www.proteinatlas.org/], and this could be at the basis both of the heterogeneous tropism of SARS-CoV-2 and of the wide spectrum of clinical pulmonary and extrapulmonary symptoms associated with the disease. ACE2 acts as the main receptor for the spike (S) protein of SARS-CoV-2; the virus entry is facilitated by the action of the transmembrane protease, serine 2 (TMPRSS2), which serves for protein S priming (i.e., the cleavage of the S protein at S1/S2 and S2' sites), thus allowing the fusion of viral and cellular membranes (Hoffmann et al., 2020).

Upon entry, the SARS-CoV-2 genome starts the production of viral proteins, including replicases: these form replication factories using membranes derived from endoplasmic reticulum (Knoops et al., 2008). Such replication factories are characterized by double-membrane vesicles that protect the double-stranded RNA transcription intermediates from detection by cytoplasmic pattern recognition receptors (PRRs), which are crucial for the initiation of innate immunity and play a key role in first-line defense (until more specific adaptive immunity is developed). The main cytoplasmic PRRs able to detect SARS-CoV-2 are thought to be MDA5 (encoded by the interferon induced with helicase C domain 1 gene, *IFIH1*) and DDX58 (encoded by the retinoic acid-inducible gene I, *RIG-I*), which initiate a signaling cascade, including interferon regulatory factor (IRF)-3 and IRF-7, to induce the transcription of type I and III interferon (IFN) genes (Thorne et al., 2021). Interferons, by autocrine and paracrine mechanisms mediated by IFN receptors (IFNARs), induce an antiviral cellular state through the expression of IFN-stimulated genes (ISGs, like *OAS1*, *OAS2*, and *OAS3*), which have direct or indirect (by attracting immune cells) antiviral functions (Schneider et al., 2014).

In parallel to the action of cytoplasmic PRRs, the virus could be recognized also by trans-membrane PRRs, like endosomal Toll-like receptors (TLRs). In particular, activation of both TLR3 and TLR7 triggers a signaling cascade that releases the main transcriptional regulator of inflammation, NF-κB, from its inhibitor (Bouayad, 2020). Upon release, NF-κB migrates to the nucleus and activates genes that code for pro-inflammatory cytokines and chemokines (e.g., *CCL2*, *CCL3*, *CCL7*, and *CXCL10*), able to recruit and activate immune cells from the bloodstream (Hariharan et al., 2021). The production of cytokines also supports the development of adaptive B and T cell responses that help clear the virus.

Severe forms of COVID-19 are associated with a dysregulation of the immune response that results in an inadequate or delayed type I IFN response (Osuchowski et al., 2021; Schultze and Aschenbrenner, 2021). In addition, SARS-CoV-2 is adept at evading innate recognition, IFN induction, and ISG-mediated pathways, by expressing a number of proteins that block these cascades (Kasuga et al., 2021). At the end, sustained hyperinflammation determines an increased immune infiltration in the lungs, a diminution in alveolar lacunar space, cell death by

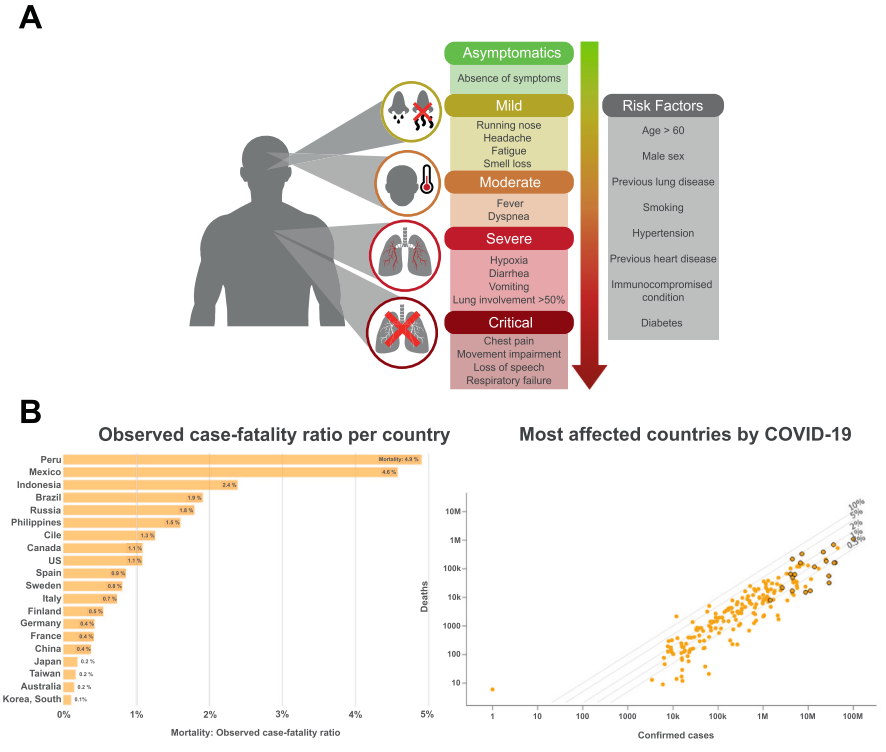


Fig. 1. COVID-19 clinical manifestations.
A. Cases infected by SARS-CoV-2 can be asymptomatic, or develop mild to critical disease. Most common symptoms are listed together with the identified risk factors.
B. Observed fatality ratios per country. On the left: fatality ratios were defined as the number of deaths divided by the number of confirmed cases; the panel shows the 20 countries most affected by COVID-19 worldwide. On the right: the diagonal lines on the chart correspond to different case fatality ratios. Countries falling on the uppermost lines have the highest observed case fatality ratios. Points with a black border correspond to the 20 most affected shown in the left panel.
Data derive from the https://coronavirus.jhu.edu/data/mortality site (maintained at the Johns Hopkins University, Coronavirus Resource center), accessed on December 28th, 2022.

apoptosis, and lung fibrosis (Merad and Martin, 2020; Rendeiro et al., 2021). Uncontrolled hyperinflammation produced by the dysregulation of the immune response can also determine the so called cytokine release syndrome (also known as cytokine storm) (Fajgenbaum and June, 2020): while the initial release of cytokines aids in eradicating the virus, excessive cytokine secretion is detrimental, as they begin to target host cells. In turn, weakened T-cell immune effects further hamper viral removal, thus starting a vicious cycle of cytokine-induced hypercytokinaemia (Tan et al., 2021). In COVID-19 patients, cytokine storm can rapidly evolve in coagulopathy, oxidative stress, and organ damage: it correlates with disease severity and has been described as a major cause of mortality (Fajgenbaum and June, 2020).

Considering the above-described SARS-CoV-2 pathophysiologic mechanisms, it is clear that severe COVID-19 can be linked, at least in part, to the patients' propensity to inflammatory injuries that affect the lungs and/or blood vessels. In turn, this tendency can be explained by susceptibility to viral infection and to immune-mediated diseases, which are both influenced by intrinsic characteristics of the virus as well as genetic factors of the host. In this Review, we aimed to recapitulate the current literature in terms of host genetic factors that specifically associate with an increased severity of the disease.

2. Genetics of COVID-19 severity

2.1. COVID-19: Study approaches and genetic consortia

At the beginning of the pandemic, the scientific community made an unprecedented effort to quickly perform genetic studies, which were mainly focused on two different phenotypes: disease severity and susceptibility to infection (The Severe Covid-19 GWAS Group, 2020; Niemi et al., 2022). In the first case, hospitalization, admission to an intensive care unit (ICU), and the use of a respiratory support (supplemental oxygen therapy only, noninvasive ventilatory support, invasive ventilatory support, or extracorporeal membrane oxygenation) were used as parameters to classify COVID-19 positive patients; here, either individuals who tested positive with non-severe disease, or subjects from the general population were used as controls. This second choice was successful in that it was possible to take advantage of the massive amount of genotype data already available through publicly available repositories. Regarding susceptibility to SARS-CoV-2 infection, this was a more arduous parameter to be defined: an ideal study should compare COVID-19 positive vs. negative individuals given exposure to the same virus strain, but this exposure is not easy to trace. Firstly because we should assign the correct SARS-CoV-2 strain to each infected, and secondly because viral exposure is influenced by a multitude of confounding factors related to human behaviors; among these: isolation periods; the use of masks, which are characterized by different levels of protection; and the type of job that can contribute to viral contacts. Again, a successful strategy in susceptibility studies was to compare true COVID-19 positives and individuals from the general population (COVID-19 Host Genetics Initiative, 2021).

Besides severity and susceptibility, additional phenotypes have been (and are currently) considered in genetic studies; among them, disease-related symptoms/outcomes (e.g., anosmia, ageusia, presence of cardiovascular events, death), long COVID-19, and vaccination response are the most commonly studied parameters, often stratifying patients according to their sex, ancestry, and age.

Considering specifically the genetic study of COVID-19 severity, different designs have been adopted. Fig. 2B summarizes these strategies: studies were aimed at identifying from one hand common single-nucleotide polymorphisms (SNPs; usually with low effect size) and from the other rare and ultra-rare variants (with higher impact) associated with different phenotypes. In the first case, strategies went from single-gene or candidate-pathway association studies, to genome-wide association studies (GWAS) and meta-analyses, up to attempts to build polygenic risk scores (PRS; also indicated as genetic risk scores, GRS). In the second case, rare variants were identified through the implementation of next-generation sequencing (NGS) approaches, like deep resequencing of panels of candidate genes, as well as whole-exome and whole-genome sequencing experiments on large cohorts of patients and controls. In many cases, these studies were performed by huge public consortia (with different participants delivering samples, freshly genotype/sequencing data, or summary statistics super-fast) and also by direct-to-consumer genetic companies. Among consortia/companies able to perform huge GWASs/meta-analyses on different populations/countries, we would like to mention:

i) The Severe Covid-19 GWAS Group: this consortium performed the first GWAS in concomitance with the peak of the pandemic in Italy and Spain in early 2020, by analyzing 1610 patients with COVID-19 severe disease (defined as respiratory failure) from 7 Italian and Spanish hospitals and 2205 controls (The Severe Covid-19 GWAS Group, 2020); the summary statistics of the first consortium analyses are available at https://ikmb.shinyapps.io/COVID-19_GWAS_Browser/. Following this seminal publication, the Severe Covid-19 GWAS Group continued in recruiting additional participants not only from Italy and Spain, but also from other European countries (Norway, Germany, Austria); this effort ended up with a second expanded GWAS meta-analysis, specifically focused on Europeans, including 3255 severe COVID-19 cases and 12,488 population controls (Degenhardt et al., 2022).

ii) The COVID-19 Host Genetics Initiative (HGI; https://www.covid19hg.org): this consortium was created to bring together the human genetics community for generating, sharing, and analyzing data to understand the genetic determinants of COVID-19 susceptibility, severity, and outcomes. The official announcement for this global initiative appeared in May 2020, again underlying the prompt response of the scientific community to the virus challenge (The COVID-19 Host Genetics Initiative, 2020). In the HGI flagship paper, dedicated to study both COVID-19 severity and susceptibility to the infection, the consortium meta-analyzed 46 studies across 19 countries comprising 6179 critically ill, 13,641 hospitalized, as well as 49,562 reported cases of SARS-CoV-2 infection, against up to 2,070,709 controls (COVID-19 Host Genetics Initiative, 2021). Following this initial publication, HGI reported an update of the analyses, enlarging the case cohorts to 9376 critically ill, 25,027 moderate or severe COVID-19 (defined as those hospitalized due to symptoms associated with the infection), and 125,584 reported cases of SARS-CoV-2 infection (with controls reaching up to 2,836,272 individuals) (COVID-19 Host Genetics Initiative, 2022). We are now waiting for a second update of the analyses, including >200,000 cases and over 3 million controls (RA, personal communication).

iii) 23andMe (https://www.23andme.com/): it is a direct-to-consumer genetic testing company that, at the beginning of the pandemic, had over 10 million microarray-based genotyped customers, of whom about 80% consented to participate in scientific researches. With these premises, 23andMe collected data from over 1 million participants, and identified 15,434 individuals who tested positive for COVID-19 infection, of whom between 636 and 1447 cases were analyzed for one of four COVID-19 severity phenotypes (i.e., hospitalization, pneumonia, respiratory support, or either respiratory support or pneumonia). Between 796,151 and 797,180 individuals were used as controls (Shelton et al., 2021).

iv) AncestryDNA (https://www.ancestry.com/dna/): this is a subsidiary of Ancestry LLC, the worldwide largest for-profit genealogy company, offering a direct-to-consumer genealogical DNA test. The company undertook a study with an approach similar to that of the 23andMe, by taking advantage of deep phenotype data

collected through a web-based survey offered to 736,723 customers in the period April–August 2020 (Roberts et al., 2022). AncestryDNA also participated to a second initiative, able to collect 52,630 COVID-19 cases and 704,016 individuals with no record of SARS-CoV-2 infection (used as controls). The other participants to this initiative were: the Geisinger Health System (GHS; https://www.geisinger.org/), the Penn Medicine BioBank (PMBB; https://pmbb.med.upenn.edu/), and the UK Biobank (UKB; https://www.ukbiobank.ac.uk/) (Horowitz et al., 2022). Both these studies deeply investigated susceptibility to infection (leading for the first time to the identification of *ACE2* as predisposing risk factor) but failed to pinpoint novel genome-wide significant loci influencing COVID-19 severity (Roberts et al., 2022; Horowitz et al., 2022).

v) The GenOMICC (Genetics of Mortality in Critical Care) initiative (https://genomicc.org/): GenOMICC is an open, collaborative, global community trying to understand and treat critical illness; born in 2015 to study emerging infections (e.g., SARS/MERS), sepsis and other forms of critical diseases, this initiative focused its efforts on 2244 COVID-19 critically ill cases. Those were recruited in 208 UK ICUs; ancestry-matched controls were selected from the UKB (5 controls to 1 case, for a total of 11,220 controls). The consortium also performed in-depth validation and replication studies, taking advantages of controls from the 100,000 Genomes and Generation Scotland initiatives for the validation step, and of cohorts from HGI and 23andMe for the replication step (The GenOMICC Investigators et al, 2021). This initial study was then followed by a second comprehensive one, in partnership with Genomics England (https://www.genomicsengland.co.uk/): here, by using whole-genome sequencing (WGS) to improve the resolution and deepen the fine-mapping of significant signals, the consortium focused on 7491 critically ill patients from 224 ICUs, and compared them to 48,400 controls (Kousathanas et al., 2022).

Table 1 summarizes the main genome-wide significant genetics findings reported by these five consortia/companies considering only the COVID-19 severe phenotype (see also paragraph 4.2.2, where we integrated the results obtained by these consortia/companies with those coming from GWAS and metanalysis performed on a population-specific perspective).

2.2. Common variant analysis

The above-reported extended researches can be considered as the luxury appetizer on the richly laid table of COVID-19 genetics. Indeed, a huge amount of studies have been performed in the last 3 years to dissect the molecular genetic basis of disease severity. To have an idea, by systematically searching PubMed until January 2nd, 2023, we retrieved a total of 25,660, 332, and 597 entries by using as keywords “COVID-19” in combination with “genetics”, “genome-wide association study”, or “polymorphisms”, respectively.

While for a minority of cases common variant analysis has been performed by extracting data from WGS outputs (Kousathanas et al., 2022), in most cases studies have been carried out by using SNP microarrays, which are now a robust, off-the-shelf, easy-to-perform industry-standard screening tool to dissect genetic predisposition (The Severe Covid-19 GWAS Group, 2020; COVID-19 Host Genetics Initiative, 2021; COVID-19 Host Genetics Initiative, 2022). These microarrays are typically designed to capture hundreds of thousands of common variants throughout the genome, but the information content of these experiments can be highly increased through imputation, a statistical strategy based on the use of publicly available, population-specific, genetic panels (e.g., those accessible through the Michigan Imputation Server, https://imputationserver.sph.umich.edu/index.html#!). Thanks to these panels, imputation allows inferencing sites of genetic variations (characterized by a population frequency $> 0.1\%$) usually up to about 10–12 millions for each individual.

2.2.1. Genome-wide studies and meta-analyses

Besides those listed in Table 1, additional GWAS and meta-analyses have been performed in the frame of COVID-19 severity, many of which aiming to dissect genetic predisposition in a population-specific perspective. This is not trivial, considering that, consistently across all disease areas, the vast majority of GWAS (91%) have been performed on people of European ancestry, whereas East Asians -the second most well-represented ethnicity- account for only 4.9% of analyzed data (Fitipaldi and Franks, 2022). GWAS and meta-analyses for COVID-19 severity do not break this rule; in addition, the study of individuals with a non-white ethnic background could help in finding novel loci, since some genetic variants that are risk factors for COVID-19 severity have high frequencies in only certain ancestries, thus becoming “measurable” and significant signals in association studies.

Table 2 summarizes the 15 novel genome-wide significant loci disclosed by population-specific GWAS/meta-analyses considering only the COVID-19 severe phenotype (see also paragraph 4.2.2, where we integrated these results with those coming from consortia/companies, listed in Table 1).

2.2.2. Bringing GWAS and meta-analysis results together

Observing the list of lead variants associated with the COVID-19 severity present in Tables 1 and 2, it clearly emerges that the effort made by the scientific community has been enormous and has led to the identification of a considerable number of significant signals, i.e. 73 unique SNP identifiers, corresponding to 41 different loci (Fig. 3A).

From these latter data, we can obviously deduce that many loci have been extensively replicated, and among them, the locus of chromosome 3p21.31 -which was the very first to be identified as associated with COVID-19 severity (The Severe Covid-19 GWAS Group, 2020)- is confirmed as the most significant. This locus corresponds to a rather complex genomic region, extremely rich in genes, and with an evident linkage disequilibrium (LD) decay going from the top associated signal (rs10490770 in HGI meta-analysis) towards centromeric positions (Fig. 3B). The region showing the highest LD is about 50-kb long, and overlaps with a haplotype block introgressed from Neanderthals; its extremely variable frequencies among populations (from $<2\%$ in East Asians and Africans, to 63% in people from Bangladesh) could indicate selection during evolution (Zeberg and Pääbo, 2020). The presence of such a strong LD has challenged scientists to unequivocally dissect the contribution of different genes, many of which are potentially implicated not only in COVID-19 severity but also in COVID-19 susceptibility. Indeed, the first HGI meta-analysis identified two independent signals at the 3p21.31 locus: the already mentioned rs10490770 (associated with severity; Table 1) and the rs22711616 lead variant (associated with infection susceptibility) (COVID-19 Host Genetics Initiative, 2021). Functional studies, omics approaches, and database mining were instrumental to focus on best candidates. In particular, the *SLC6A20* gene, coding for a sodium imino-acid transporter, has been proposed as the causal gene for infection susceptibility, considering both its functional interaction with the SARS-CoV-2 receptor *ACE2* (Vuille-dit-Bille et al., 2015) and its expression in epithelial alveolar cells (https://www.proteinatlas.org/). As for severity, the leucine zipper transcription factor like 1 (*LZTFL1*) gene has been in-depth investigated: chromosome-conformation capture and gene-expression analysis evidenced that the gain-of-function risk A allele of the SNP rs17713054 significantly affects an enhancer that upregulates *LZTFL1*, which in turn modulates the viral response pathway termed epithelial-mesenchymal transition (EMT) in pulmonary epithelial cells (Downes et al., 2021). Apart from *LZTFL1*, the 3p21.31 region also harbors a number of chemokine receptor genes, biologically plausible for having a role in COVID-19 severity: i) *CCR9* (CC motif chemokine receptor 9, a key regulator in early respiratory allergic inflammation) (López-Pacheco et al., 2016); ii) *CXCR6* (C-X

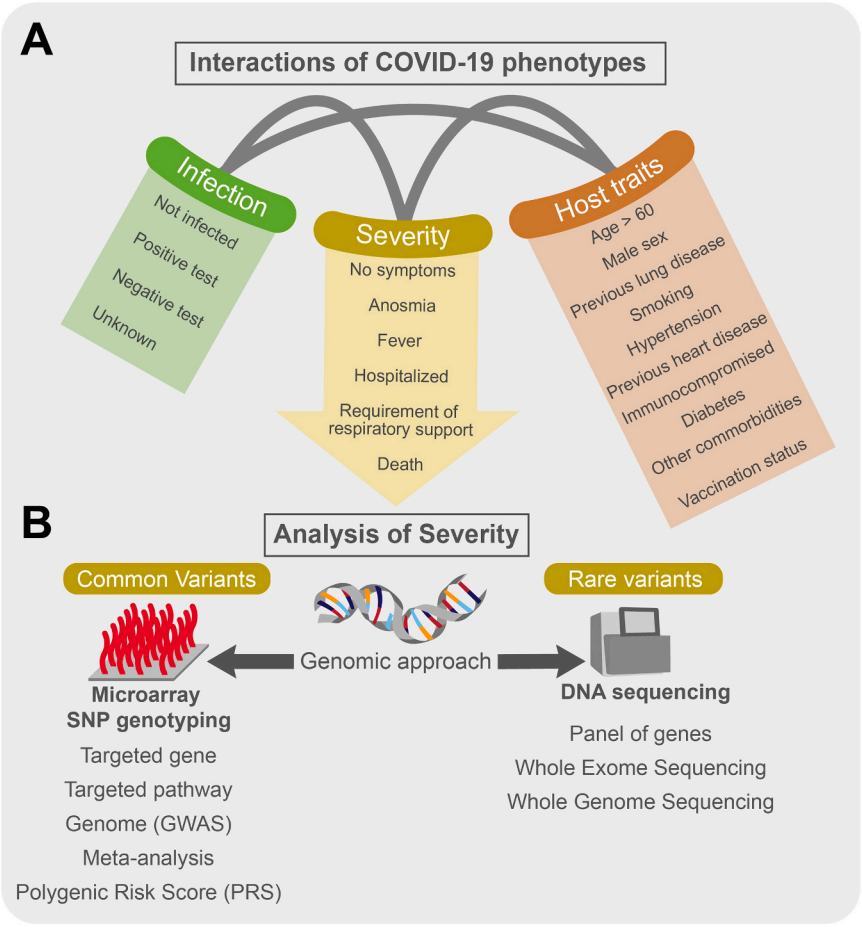


Fig. 2. Different approaches to the study of COVID-19 genetics.
A. Genetic studies on COVID-19 have focused mainly on two phenotypes: disease severity and susceptibility to infection. Alternative phenotypes (listed) and interaction analyses have tried to deepen the knowledge on host genetics.
B. Summary of strategies used to unravel the genetic basis of COVID-19 severity.

Table 1
Genome-wide significant results of GWASs and meta-analyses on COVID-19 severity performed by public consortia or companies on patients/controls belonging to different populations.

Consortium/Company	*Cases * (description)*	*Controls * (description)*	*Population/ Country*	*Lead variant*	*Location ***	*P value [OR (95% CI)]*	*Candidate causal gene/s*	*Reference*
The Severe Covid-19 GWAS Group	1610 (COVID-19 positive, hospitalized with respiratory support)	2205 (population controls, unknown COVID-19 status)	Italy, Spain	rs11385942	chr3:45834968	1.15×10^{-10} [1.77 (1.48–2.11)]	*SLC6A20, LZTFL1, FYCO1, CXCR6, XCR1, CCR9*	(The Severe Covid-19 GWAS Group, 2020)
				rs657152	chr9:133263762	4.95×10^{-8} [1.32 (1.20–1.47)]	*ABO*	
				rs35731912	chr3:45848457	2.32×10^{-21} [1.78 (1.72–1.85)]	*LZTFL1*	
	3255 (hospitalized with respiratory support)	12,488 (population controls)	Italy, Spain, Norway, Germany, Austria	rs11085725	chr19:10351837	1.02×10^{-10} [1.27 (1.25–1.28)]	*TYK2*	(Degenhardt et al., 2022)
				rs687289	chr9:133261703	4.49×10^{-10} [1.24 (1.22–1.26)]	*ABO*	
				rs12610495	chr19:4717660	2.87×10^{-8} [1.29 (1.26–1.32)]	*DPP9*	
	6526 (hospitalized with respiratory support + B2:	719,075 (population controls)	Italy, Spain, Norway, Germany, Austria,	**rs1819040**	chr17:46142465	3.27×10^{-11} [0.88 (0.84–0.92)]	*KANSL1*	

(*continued on next page*)

Table 1 (*continued*)

Consortium/Company	*Cases * (description)*	*Controls * (description)*	*Population/ Country*	*Lead variant*	*Location **	*P value [OR (95% CI)]*	*Candidate causal gene/s*	*Reference*
	Hospitalized COVID19+)		plus HGI B2 cohort					
	14,467 (critically ill: hospitalized with mechanical ventilation + A2: Critically ill COVID19+)	1,306,293 (population controls)	Italy, Spain, Norway, Germany, Austria, plus HGI A2 cohort	**rs1405655**	chr19:50379362	3.25×10^{-8} [1.09 (1.06–1.13)]	*NAPSA*	
The COVID-19 Host Genetics Initiative	13,641 (COVID-19 positive and hospitalization)	2,070,709 (population controls, unknown COVID-19 status)	European, Admixed American, African, Middle Eastern, South Asian, East Asian (19 countries)	rs10490770	chr3:45823240	1.44×10^{-73} [1.65 (1.56–1.74)]	*LZTFL1*	(COVID-19 Host Genetics Initiative, 2021)
				rs1886814	chr6:41534945	1.11×10^{-9} [1.26 (1.17–1.36)]	*FOXP4*	
				rs72711165	chr8:124324323	2.13×10^{-9} [1.37 (1.24–1.52)]	*TMEM65*	
				rs10774671	chr12:112919388	6.14×10^{-10} [1.11 (1.07–1.14)]	*OAS1*, *OAS3*, *OAS2*	
				rs1819040	chr17:46142465	1.83×10^{-10} [0.88 (0.84–0.91)]	*ARHGAP27*, *PLEKHM1*, *LINC02210-CRHR1*, *CRHR1*, *SPPL2C*, *MAPT*, *STH*, *KANSL1*, *LRRC37A*, *ARL17B*, *LRRC37A2*, *ARL17A*, *NSF*, *WNT3*	
				rs77534576	chr17:49863303	2.26×10^{-07} *** [1.26 (1.15–1.37)]	*KAT7*, *TAC4*	
				rs2109069	chr19:4719431	2.76×10^{-17} [1.15 (1.12–1.19)]	*DPP9*	
				rs74956615	chr19:10317045	5.05×10^{-10} [1.27 (1.18–1.36)]	*ICAM1*, *ICAM4*, *ICAM5*, *ZGLP1*, *FDX2*, *RAVER1*, *ICAM3*, *TYK2*	
				rs13050728	chr21:33242905	2.72×10^{-20} [0.86 (0.83–0.89)]	*IFNAR2*	
	25,027 (moderate or severe COVID-19, defined as those hospitalized due to symptoms associated with the infection)	2,836,272 (control individuals)	Europeans, Admixed Americans, Africans, Middle Eastern, South Asians, East Asians (25 countries)	rs67579710	chr1:155203736	1.76×10^{-10} [0.9 (0.87–0.93)]	*EFNA1*, *SLC50A1*, *DPM3*, *KRTCAP2*, *TRIM46*, *MUC1*, *THBS3*, *MTX1*, *GBA*, *FAM189B*, *SCAMP3*, *CLK2*, *HCN3*, *PKLR*, *FDPS*, *RUSC1*, *ASH1L*, *MSTO1*	(COVID-19 Host Genetics Initiative, 2022)
				rs35508621	chr3:45838989	4.81×10^{-100} [1.49 (1.44–1.55)]	*LZTFL1*	
				rs111837807	chr6:31153455			

(*continued on next page*)

Table 1 (*continued*)

Consortium/Company	*Cases * (description)*	*Controls * (description)*	*Population/ Country*	*Lead variant*	*Location **	*P value [OR (95% CI)]*	*Candidate causal gene/s*	*Reference*
						2.33×10^{-11} [1.13 (1.09–1.17)]	*C6orf15, PSORS1C1, CDSN, PSORS1C2, CCHCR1, TCF19, POU5F1, HLA-C, HLA-B*	
				rs41435745	chr6:41522644	6.85×10^{-13} [1.2 (1.14–1.27)]	*FOXP4*	
				rs721917	chr10:79946568	1.69×10^{-8} [1.06 (1.04–1.08)]	*SFTPD*	
				rs35705950	chr11:1219991	6.47×10^{-9} [0.89 (0.86–0.93)]	*MUC5B*	
				rs766826	chr11:34507219	$2.93 \times 10–^{12}$ [0.92 (0.9–0.94)]	*ELF5*	
				rs10774679	chr12:112936943	6.97×10^{-13} [1.08 (1.06–1.1)]	*OAS1, OAS3, OAS2*	
				rs12809318	chr12:132564254	3.22×10^{-9} [0.94 (0.92–0.96)]	*FBRSL1*	
				rs117169628	chr16:89196249	2.55×10^{-8} [1.09 (1.06–1.13)]	*ACSF3, CDH15, SLC22A31*	
				rs61667602	chr17:45707983	3.77×10^{-12} [0.91 (0.89–0.94)]	*ARHGAP27, PLEKHM1, LINC02210-CRHR1, CRHR1, SPPL2C, MAPT, STH, KANSL1, LRRC37A, ARL17B, LRRC37A2, ARL17A, NSF*	
				rs77534576	chr17:49863303	4.29×10^{-11} [1.23 (1.16–1.31)]	*KAT7, TAC4*	
				rs2109069	chr19:4719431	2.05×10^{-22} [1.12 (1.09–1.14)]	*DPP9*	
				rs11085727	chr19:10355447	2.66×10^{-14} [1.09 (1.07–1.12)]	*TYK2*	
				rs1405655	chr19:50379362	5.34×10^{-11} [1.07 (1.05–1.1)]	*NAPSA, NR1H2, POLD1*	
				rs13050728	chr21:33242905	5.13×10^{-23} [0.9 (0.88–0.92)]	*IFNAR2*	
23andMe	1447 (COVID-19 positive, and respiratory support or pneumonia)	796,151 (General population with unknown COVID-19 status)	Europeans, Latinos, African Americans (all mainly from US)	rs13078854	chr3:45820340	1.6×10^{-18} [0.59 (0.53–0.67)]	*LZTFL1*	(Shelton et al., 2021)
AncestryDNA	2184 (COVID-19	45,185 (COVID-19 positive and	Africans, East Asians, Europeans,	**rs73064425**	chr3:45859597	2.2×10^{-18} [1.58 (1.43–1.75)]	*LZTFL1*	(Horowitz et al., 2022)

(*continued on next page*)

Table 1 (*continued*)

Consortium/Company	*Cases * (description)*	*Controls * (description)*	*Population/ Country*	*Lead variant*	*Location **	*P value [OR (95% CI)]*	*Candidate causal gene/s*	*Reference*
	positive and severe)	not hospitalized)	Latin Americans, South Asians					
The GenOMICC (Genetics of Mortality in Critical Care) initiative	2244 (critically ill patients with COVID-19 from ICUs; the primary analysis on data from 1676 individuals of European ancestry) ****	8380 (controls from population genetic studies in the UK; those used in the primary analyses) ****	Europeans, South Asians, Africans, East Asians	**rs73064425**	chr3:45859697	3.6×10^{-32} [2.1 (1.88–2.45)]	*LZTFL1*	(The GenOMICC Investigators et al., 2021)
				rs9380142	chr6:29831017	1.8×10^{-8} [1.3 (1.18–1.43)]	*HLA-G*	
				rs143334143	chr6:31153649	2.6×10^{-24} [1.8 (1.61–2.13)]	*CCHCR1*	
				rs3131294	chr6: 32212369	1.3×10^{-10} [1.5 (1.28–1.66)]	*NOTCH4*	
				rs10735079	chr12:112942203	2.8×10^{-9} [1.3 (1.18–1.42)]	*OAS1*, *OAS3*	
				rs2109069	chr19:4719431	4.0×10^{-12} [1.4 (1.25–1.48)]	*DPP9*	
				rs74956615	chr19:10317045	2.2×10^{-13} [1.6 (1.35–1.87)]	*TYK2*	
				rs2236757	chr21:33252612	5.0×10^{-8} [1.3 (1.17–1.41)]	*IFNAR2*	
	7491 (critically ill patients from ICUs)	48,400 (controls were from the 100,000 Genomes Project cohort, $n = 46{,}770$; and mild COVID-19, $n = 1630$)		rs114301457	chr1:155066988	6.8×10^{-10} [2.4 (1.82–3.14)]	*EFNA4*	(Kousathanas et al., 2022)
				rs7528026	chr1:155175305	7.16×10^{-9} [1.4 (1.24–1.55)]	*TRIM46*	
				rs41264915	chr1:155197995	1.02×10^{-12} [1.3 (1.19–1.37)]	*THBS3*	
				rs1123573	chr2: 60480453	9.85×10^{-10} [1.1 (1.09–1.18)]	*BCL11A*	
				rs2271616	chr3:45796521	9.9×10^{-17} [1.3 (1.21–1.37)]	*SLC6A20*	
				rs73064425	chr3:45859597	1.97×10^{-133} [2.7 (2.51–2.94)]	*LZTFL1*	
				rs343320	chr3:146517122	4.94×10^{-9} [1.2 (1.16–1.35)]	*PLSCR1*	
				rs56162149	chr5:131995059	7.65×10^{-11} [1.2 (1.13–1.26)]	*ACSL6*	
				rs9271609	chr6:32623820	3.26×10^{-9} [1.1 (1.09–1.19)]	*HLA-DRB1*	
				rs2496644	chr6:41515007	7.59×10^{-15} [1.4 (1.32–1.60)]	*LINC01276*	
				rs28368148	chr9:21206606	1.93×10^{-9} [1.7 (1.45–2.09)]	*IFNA10*	
				rs61882275	chr11:34482745	1.61×10^{-10} [1.1 (1.10–1.20)]	*ELF5*	
				rs56106917	chr12:132489230	2.08×10^{-9} [1.1 (1.09–1.18)]	*FBRSL1*	
				rs9577175	chr13:112889041		*ATP11A*	

(*continued on next page*)

Table 1 (*continued*)

Consortium/Company	*Cases * (description)*	*Controls * (description)*	*Population/ Country*	*Lead variant*	*Location **	*P value [OR (95% CI)]*	*Candidate causal gene/s*	*Reference*
						3.71×10^{-11} [1.2 (1.12–1.24)]		
				rs4424872	chr15:93046840	8.61×10^{-13} [2.4 (1.87–3.01)]	*RGMA*	
				rs117169628	chr16:89196249	4.4×10^{-9} [1.2 (1.12–1.26)]	*SLC22A31*	
				rs2532300	chr17:46152620	4.19×10^{-9} [1.2 (1.10–1.22)]	*KANSL1*	
				rs3848456	chr17:49863260	4.19×10^{-11} [1.5 (1.33–1.70)]	–	
				rs12610495	chr19:4717660	3.91×10^{-36} [1.3 (1.27–1.38)]	*DPP9*	
				rs73510898	chr19:10305768	1.57×10^{-11} [1.3 (1.19–1.37)]	*ZGLP1*	
				rs34536443	chr19:10352442	6.98×10^{-17} [1.5 (1.36–1.65)]	*TYK2*	
				rs368565	chr19: 48697960	3.55×10^{-11} [1.1 (1.1–1.2)]	*FUT2*	
				rs17860115	chr21:33230000	9.69×10^{-22} [1.2 (1.19–1.3)]	*IFNAR2*	
				rs8178521	chr21:33287378	3.53×10^{-12} [1.2 (1.12–1.23)]	*IL10RB*	
				rs35370143	chr21:33959662	1.24×10^{-9} [1.3 (1.17–1.36)]	*LINC00649*	

Only genome-wide significant results ($<5 \times 10^{-8}$) are listed. Lead variants that have been replicated at least once at a $P < 5 \times 10^{-8}$ are bolded; genes belonging to regions pinpointed by genome-wide significant association signals and replicated at least once are underlined.
* Surviving all quality-control checks in the corresponding study; ** According to UCSC Genome Browser on Human (GRCh38/hg38); *** This P value becomes genome-wide significant (4.37×10^{-9}) when considering critical COVID-19; **** In this study, different control cohorts were used for replication purposes: the best obtained *p* value is reported for each analyzed SNP.
CI, Confidence interval; OR, Odds ratio.

motif chemokine receptor 6, which recruits CD8-resident memory T cells in the airways against respiratory pathogens) (Wein et al., 2019); and iii) *CCR2* (C—C chemokine receptor type 2, a proinflammatory molecule, whose expression is increased in carriers of 3p21.31 risk alleles) (The GenOMICC Investigators et al, 2021).

Besides the 3p21.31 region, a second locus has been widely replicated in GWAS: the *ABO* histo-blood group gene. This gene was initially associated with respiratory failure, in the very first GWAS performed on COVID-19 phenotypes (The Severe Covid-19 GWAS Group, 2020); later on, it was cause of debate, until its consistent and robust association with infection susceptibility was demonstrated (COVID-19 Host Genetics Initiative, 2022; Shelton et al., 2021; COVID-19 Host Genetics Initiative et al., 2021).

More in general, as evidenced by HGI, the identified loci well recapitulate the history of SARS-CoV-2 viral infection, with a number of relevant genes that can be linked to three key pathways involved in the pathogenesis of the disease, i.e. viral entry/defense, type I interferon response, and maintenance of healthy lung tissue (COVID-19 Host Genetics Initiative, 2022). We confirmed these results by performing an ad-hoc enrichment analysis, using as starting dataset the non-redundant list of lead variants presented in Tables 1 and 2. This list was submitted to the web-based variant annotation tool SNPnexus (https://www.snp-nexus.org/v4/) (Oscanoa et al., 2020), which evidenced: i) the occurrence of the association signals within/in proximity of a total of 171 refseq genes (of which 121 coding and 20 non-coding) (Fig. 3C); ii) in the vast majority of cases, lead variants map within introns, whereas 10 are predicted to affect the coding sequence; iii) splice-site and non-synonymous SNPs are predicted to be deleterious -with high confidence- in 9 cases (SIFT-based prediction; Fig. 3D); and iv) the Reactome Pathways Enrichment analyses disclosed the “Immune system” as the major field involved by COVID-19-related genetic variation, with a preference for “Signaling by Interleukins” (interleukins 4, 10, 13, and

Table 2
Novel genome-wide significant results of GWASs and meta-analyses on COVID-19 severity performed on patients/controls belonging to specific populations.

Population	*Cases*	*Controls*	*Lead variant*	*Location* *	*P value [OR (95%CI)]*	*Candidate causal gene/s*	*Reference*
Serbian	Moderate ($n = 46$) and severe ($n = 34$) COVID-19 patients	Mild ($n = 48$) COVID-19 patients	rs61964606	chr13:70189032	1.91×10^{-8} [10.12 (4.46–22.95)]	*KLHL1, ATXN8, ATXN80S*	(Zecevic et al., 2022)
Japanese	440 (severe hospitalized cases, <65 years of age)	2377 (population based controls, unknown COVID-19 status, <65 years of age)	rs60200309	chr5:170092608	1.2×10^{-8} [2.01 (1.58–2.55)]	*DOCK2*	(Namkoong et al., 2022)
Spanish	7344 (hospitalized patients)	7011 (COVID-19 positive cases and 1068 population controls)	rs10813976	chr9:33426577	2.7×10^{-8} [0.18–0.03]**	*AQP3*	(Cruz et al., 2022)
			rs77127536	chr19:35687796	1.3×10^{-8} [−0.22–0.04] **	*UPK1A, ZTBT32*	
Brazilian	3533 (COVID-19 positive and hospitalized; all from Sao Paulo, Brazil)	1700 (COVID-19 positive and not hospitalized)	rs11240388	chr1:205208489	3.99×10^{-8} [1.35]***	*RBBP5, DSTYK, TMCC2*	(Pereira et al., 2022)
Indian	96 (COVID-19 deceased patients)	148 (asymptomatic patients)	rs34279101	chr14:76832815	4.12×10^{-8} n.r.	*ANGEL1, VASH1-AS1, AC007376.2, RN7SKP17, misc_RNA, AF111169.1, LRRC74A, RPL22P2, AF111169.4*	(Pandit et al., 2022)
Chinese ****	632 (critical patients)	3021 (healthy controls)	rs7598285	chr2:6365431	1.69×10^{-10} [0.65]***	*LINC01247*	(Gong et al., 2022)
			rs10519086	chr2:67035514	6.82×10^{-11} [0.42]***	*LINC01799, LINC01828*	
			rs72809129	chr2:77248537	1.42×10^{-9} [0.67]***	*LRRTM4*	
			rs7422259	chr2:166521485	3.60×10^{-10} [0.49]***	*SCN7A, XIRP2*	
			rs2069837	chr7:22728408	4.64×10^{-16} [0.49]***	*IL6*	
			rs17158686	chr7:84165440	4.99×10^{-8} [0.61]***	*SEMA3A*	
Chinese	474 (Critically ill or other severe conditions)	1615 (COVID-19 positive without severe symptoms + population-based controls with unknown COVID-19 status)	rs74490654	chr19:19163581	1.22×10^{-8} [8.73 (4.14–18.41)] *****	*MEF2B*	(Wu et al., 2021)
Chinese	885 (severe or critical COVID-19 patients)	546 (mild or moderate COVID-19 patients)	rs1712779	chr11:114726432	1.38×10^{-8} [0.49 (0.39–0.63)]	*NXPE2*	(Li et al., 2021)
			rs10831496	chr11:88824823	4.04×10^{-8} [1.66 (1.38–1.98)]	*GRM5*	

Only genome-wide significant results ($<5 \times 10^{-8}$) are listed. * According to UCSC Genome Browser on Human (GRCh38/hg38); ** In this study, beta and SE were provided and are listed in this order; *** Confidence interval not reported; **** For this study, results of the meta-analysis surviving the genome-wide threshold are reported (437 cases 2551 controls in the discovery cohort, 195 cases 470 controls in the replication cohort); ***** These values were calculated through a meta-analysis with the HGI (B2_release3) cohort (accounting for 3199 cases and 897,488 controls); n.r., not reported.

20) and "Interferon signaling" domains (Fig. 4A). Fig. 4B shows the interactions among proteins tagged by lead variants (the list of non-redundant proteins was submitted to String, a database of known and predicted protein-protein interactions; https://string-db.org/; (Szklarczyk et al., 2021)).

2.2.3. Candidate-gene and pathway association studies

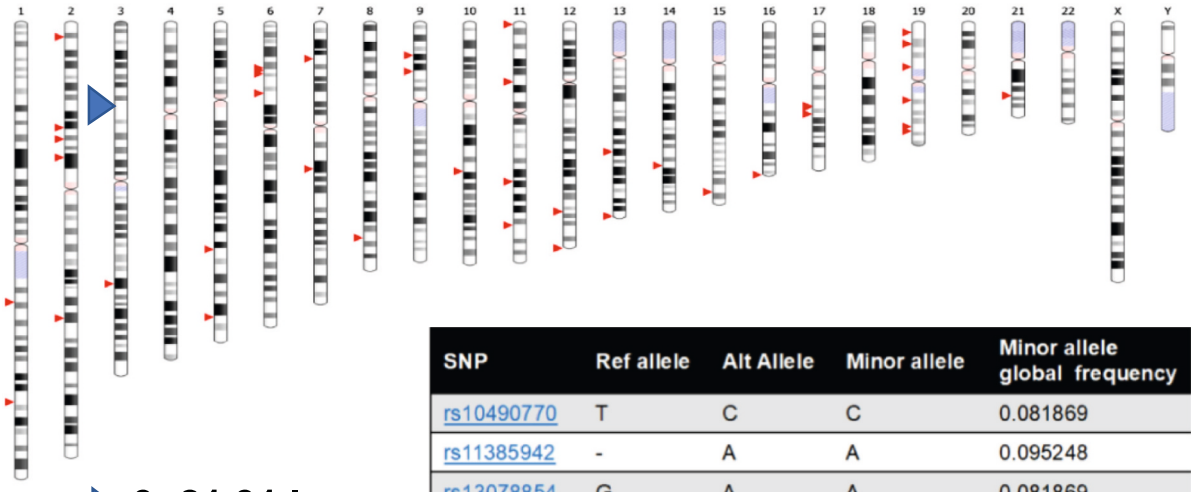
Normally, candidate gene studies stem on the investigation of variants involving genes/pathways that are known to be critical in a disease or trait development. Severe COVID-19 does not break this rule. In this respect, association analyses have focused initially on "obvious" candidates, such as *ACE2* and *TMPRSS2* (for susceptibility to infection), on genes of the innate and adaptive immunity (for disease severity), as well as on genes of the human leukocyte antigen (HLA) system (for both disease severity and susceptibility to infection). Table 3 shows the list of gene/pathway-candidate association studies on disease severity performed on at least 100 patients and reporting significant results with a P value ≤ 0.01. We searched the PubMed repository for original articles that analyzed potential associations between genetic polymorphisms and severity for COVID-19, up to August 2022.

Importantly, we did not include among candidate genes those of HLA, which were in-depth investigated by dozens of research groups, often with conflicting results. For a specific focus on this topic, we suggest a series of excellent reviews in this field (Dieter et al., 2022; Ishak et al., 2022).

We found a total of 14 studies, reporting 26 signals at $P \leq 0.01$ in 17 different genes. The strongest signal of association was found for the e4e4 haplotype characterizing the *APOE* gene ($P = 1.19 \times 10^{-6}$, OR = 2.31, 95%CI = 1.65–3.24) (Dite et al., 2021), and not surprisingly it showed up in the study with the highest number of analyzed individuals among those reported in Table 3 (it included >322 thousands individuals from the UK Biobank). The second best signal was found -again- for an haplotype, composed of SNPs rs10824844 and rs10824845, mapping in proximity of the *MBL2* gene ($P = 1.04 \times 10^{-5}$, OR = 1.88, 95%CI = 1.44–2.45) (Stravalaci et al., 2022). The association with this gene was subsequently replicated in two additional works (Asselta et al., 2022; Medetalibeyoglu et al., 2021), making this gene the most replicated among those not reaching the genome-wide significance threshold.

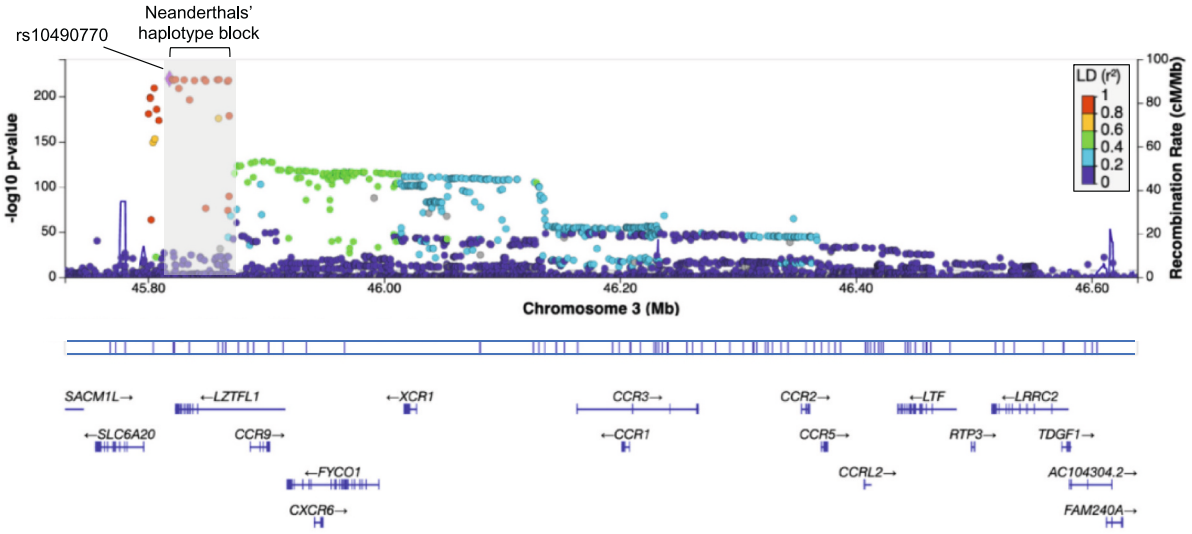
Concerning overall these candidate-gene/pathway association

A

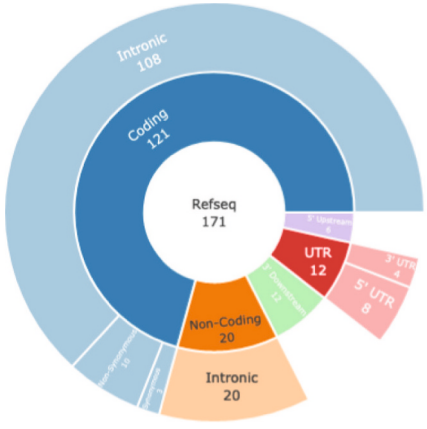


SNP	Ref allele	Alt Allele	Minor allele	Minor allele global frequency
rs10490770	T	C	C	0.081869
rs11385942	-	A	A	0.095248
rs13078854	G	A	A	0.081869
rs2271616	G	T	T	0.080871

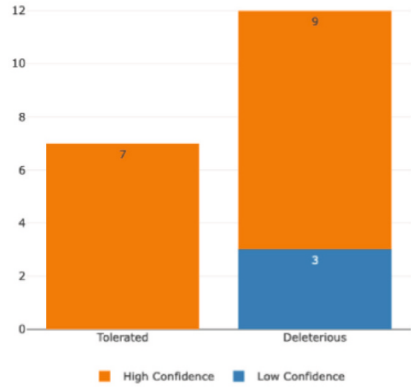
B



C



D



(caption on next page)

Fig. 3. GWAS on common variants: summary of the top results.
A. Karyotype plot of lead variants. The panel shows the distribution at the chromosome level of all genome-wide significant signals found in GWAS studies (listed in Tables 1 and 2). Loci are indicated with a red arrowhead, with the exception of the 3p21.31 locus, which is pointed out by a blue arrowhead. As for this locus, the table shows the 4 SNPs found in different GWAS.
B. Regional plot for associations in the 3p21.31 region. The *P* values of the SNPs in this region refer to HGI analysis A2, release 7. The rs10490770 top SNP is shown as a purple diamond; all LD values were referred to this lead SNP. The 50-kb region evidenced with a vertical band in light grey correspond to the haplotype block introgressed from Neanderthals.
C. and **D.** Mapping genome-wide significant signals. Panel C displays the distribution of the lead association signals with respect to refseq genes, whereas panel D shows the potential pathogenic impact of those signals mapping within coding genes.
Panels A, C, and D were produced using the web-based variant annotation tool SNPnexus (https://www.snp-nexus.org/v4/) (Oscanoa et al., 2020). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

studies, it is interesting to underline that: i) only single populations have been considered, but in many instances these studies deal with so-far "neglected" ethnic groups (i.e., Egyptian, Saudi, Indian, and Iranian populations); ii) two genes, well known to be implicated in susceptibility to infection, have been associated also with disease severity (i.e., *ACE2* and *TMPRSS2*); and iii) the lower statistical power offered by the analyzed cohorts was "compensated" with the use of methodological approaches usually not considered in GWAS and meta-analyses (i.e., genotypic association analysis, haplotype analysis, as well as PRS based on few candidates).

2.3. Rare variant analysis

Rare variant studies for COVID-19 severity are still in their infancy, as also testified by PubMed searches (performed on January 2nd, 2023): we retrieved 35 and 80 hits by using as keywords "COVID-19" in combination with "rare variants" and "exome sequencing", respectively. Of note, though also the genome-sequencing approach was used to investigate the impact of rare variants on COVID-19 severity, a specific PubMed search using "genome sequencing" as keyword is misleading in this case, due to the high number of studies investigating the virus genome rather than the host genome.

Generally speaking, rare genetic defects are expected to show larger effect sizes, because of evolutionary pressure on highly deleterious variants, thus providing unique insights into genetic predisposition to COVID-19 severity. In this frame, the first pivotal study appearing in the literature on rare variants reported on 659 patients with severe COVID-19 and 534 subjects with asymptomatic infection: Zhang and colleagues demonstrated that 3.5% of severe patients had genetic defects at 8 of the 13 loci involved in inborn errors of type I IFN immunity (loss-of-function mutations were found by next-generation sequencing in the *IRF3*, *IRF7*, *IFNAR1*, *IFNAR2*, *TLR3*, *UNC93B1*, *TICAM1*, and *TBK1* genes; $P = 0.01$) (Zhang et al., 2020). A nice corroboration of these results was the observation that a B-cell autoimmune phenocopy of inborn errors of type I IFN immunity accounts for life-threatening COVID-19 pneumonia in 2.6% of women and 12.5% of men, thus providing a first explanation for the excess of men among patients with severe disorder (Bastard et al., 2020). These findings, however, were not replicated in a series of well-powered subsequent works (Kousathanas et al., 2022; Povysil et al., 2021; Kosmicki et al., 2021). In particular: i) Kousathanas and colleagues analyzed 7491 individuals with critical COVID-19 and 5391 controls by WGS and not only did not find any significant association among the 13 genes of the type I IFN pathway (burden test for each gene showed $P > 0.05$), but neither found significant associations at a gene-wide-significance level (with all individual rare variants included in the tests showing $P > 10^{-5}$) (Kousathanas et al., 2022); ii) Povysil et al. performed WES or WGS of 713 severe COVID-19 cases and 15,033 ancestry-matched controls, and identified just one rare loss-of-function mutation in one of the 13 type I IFN genes among severe cases (Povysil et al., 2021); and iii) Kosmicki and colleagues examined exome data from 586,157 individuals, including 20,952 with COVID-19, and did not identify any clear association with rare variants either exome wide, or specifically analyzing the 13 IFN genes, the genes located in susceptibility loci identified by HGI, as well additional candidates with immunologic relevance and/or therapeutic potential (Kosmicki et al., 2021).

Those listed above are not the only studies investigating rare variants in the COVID-19 population; however, results coming from this type of analyses have been overall disappointing. So far, the only notable exception concerns the SARS-CoV-2 sensor toll-like receptor *TLR7* gene (on chromosome X), which was associated with COVID-19 severity multiple times (van der Made et al., 2020; Fallerini et al., 2021; Asano et al., 2021; Butler-Laporte et al., 2022). In particular, the very recent study by Butler-Laporte et al. (Butler-Laporte et al., 2022) reported on the analyses of WES/WGS data from 21 cohorts across 12 countries (accounting for a total of 5085 severe disease cases and 571,737 controls), and found an association signal consistent across sexes corresponding to an OR of 5.3 for severe disease (95%CI = 2.8–10.1, $P = 5.4 \times 10^{-7}$). Nonetheless, larger studies are still necessary to provide additional insights on rare variants influencing COVID-19 phenotypes.

3. Discussion, conclusions and future perspectives

The world is still suffering from the COVID-19 pandemic and the related global economic, social, and political troubles are poised to worsen in the future. Notwithstanding the intensive efforts of scientists, we are still in the need of better understanding the causal relationships between host genetics and COVID-19, in order to identify biomarkers for individuals at risk, as well as to provide potential targets for therapy.

The initial GWAS was indeed a milestone in directing many subsequent analyses: the choice to focus on severity in a "simplified" manner (basically classifying patients on the basis of the respiratory support) was successful in immediately giving answers on the contribution of genetic variants to the disorder (The Severe Covid-19 GWAS Group, 2020). Many following studies remained sticked to this strategy, and involved from few hundreds of cases/controls to huge cohorts collected world-wide (Tables 1-3). On the other hand, many attempts have been made to better dissect genetic signals involved in the pathogenesis/modulation of the phenotype. A paradigmatic example can be considered the study performed by Nakanishi and colleagues, who analyzed 7185 hospitalized patients to better examine the association of the major common COVID-19 genetic risk factor (the chromosome 3 locus, tagged by the rs10490770 SNP) with mortality, COVID-19-related complications, and laboratory values (Nakanishi et al., 2021). The most interesting result pertained the age-based stratification analysis: rs10490770 was associated with death or severe respiratory failure especially in young patients [if <60 years OR = 2.7 (95%CI = 1.8–3.9); if >60 years, OR = 1.5 (95%CI = 1.2–1.8)], with individuals of 60 years and younger who died or experienced severe respiratory failure carrying the risk variant in 32.3% of cases (compared with 13.9% of those not experiencing these outcomes) (Nakanishi et al., 2021).

Additional efforts pertained the definition of genetic profiles predisposing to severe COVID-19. By encompassing the combined effect of different loci, the definition of a GRS should indeed represent a better tool to predict disease risk than single loci. Some initial attempts have been performed by considering a low number of SNPs in the GRS model (usually only including the best associated signals published in the literature). Overall, these attempts show that common variants define a

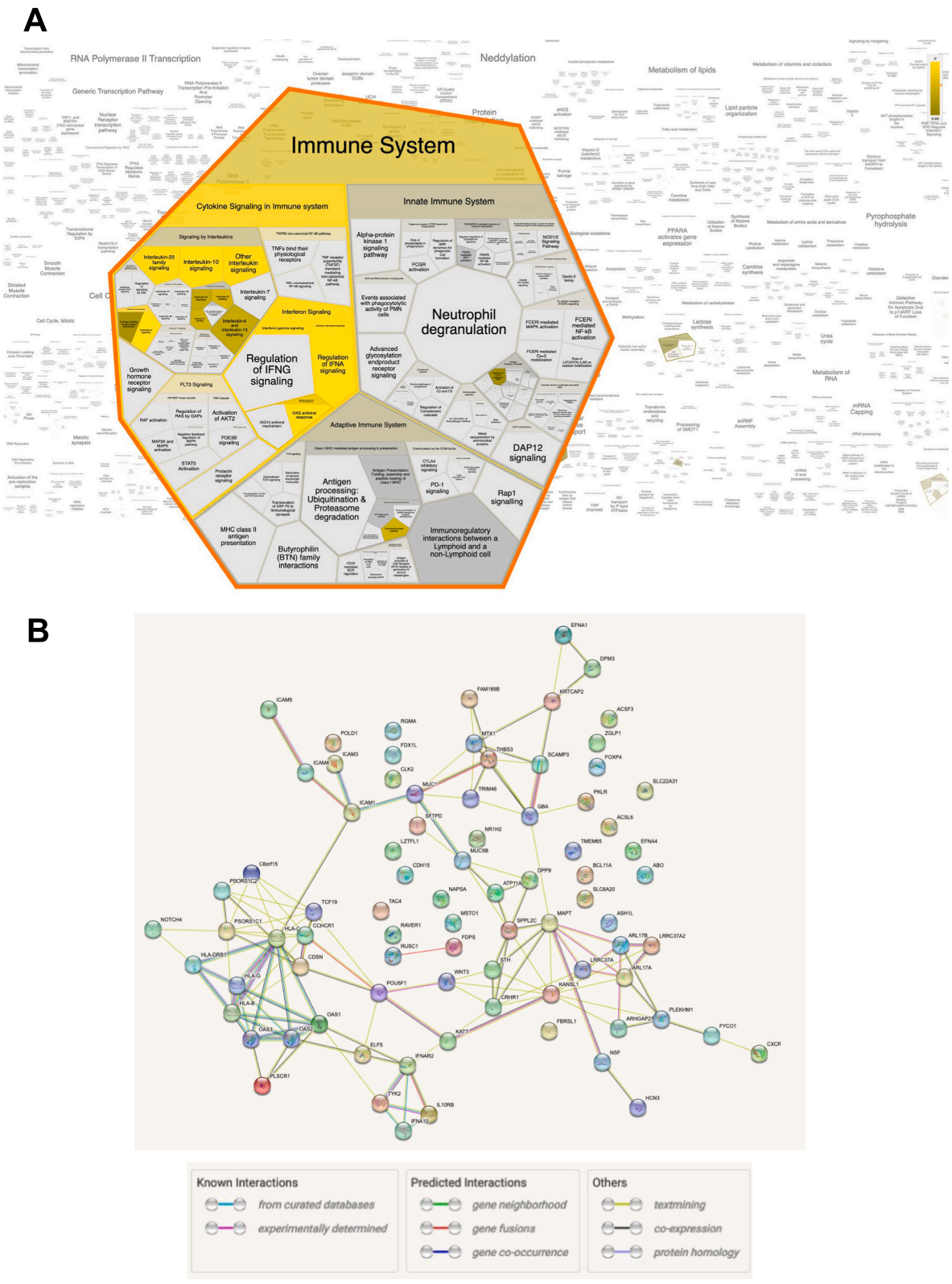


Fig. 4. Enrichment and String analyses for genome-wide significant signals.
A. Pathways enrichment analysis. The analysis was performed using the Reactome database; enriched pathways are evidenced with different shades of yellow. The panel was produced using the web-based variant annotation tool SNPnexus (https://www.snp-nexus.org/v4/) (Oscanoa et al., 2020).
B. String analysis. Interactions among proteins tagged by lead variants are indicated. The figure was prepared using String, the database of known and predicted protein-protein interactions (https://string-db.org/; (Szklarczyk et al., 2021)). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Gene/pathway association studies for COVID-19 severity.

Population	Analyzed cohort	Gene/pathway	Associated top hit	*P value* *[OR (95%CI)]* *	Reference
Egyptian	141 patients 100 healthy controls	*IFN-λ*, *TLL1*, *DDR1*	*DDR1* rs4618569	0.007	(Agwa et al., 2021)
Saudi	880 patients analyzed for mortality	*IFITM3*	rs12252	0.01 [2.20 (1.16–4.20)]	(Alghamdi et al., 2021)
British	1018 severe patients 564 without severe disease	64 SNPs identified as associated from published data	Score including 64 SNPs	<0.001 [1.19 (1.15–1.22)]	(Dite et al., 2021)
British	622 cases 322,326 controls	*APOE*	e4e4 haplotype	1.19×10^{-6} [2.31 (1.65–3.24)]	(Kuo et al., 2020)
German	297 severe cases 253 controls	*ACE1*, *ACE2*	*ACE2* rs2285666	0.002 [3.04 (1.47–6.27)]	(Möhlendick et al., 2021)
Indian	149 mild cases 120 severe cases	*ACE1*	*DD ACE1* genotype	0.002 [3.69 (1.61–8.43)]	(Verma et al., 2021)
Austrian	361 cases (of which 190 hospitalized and 70 in ICU) 260 controls	*HLA-E, KLRC2*	*NKG2Cdel* *HLA-E*0101*	0.0006 [OR = 2.6 for hospitalization] & <0.0001 [OR = 7.1 for ICU] 0.01 [OR = 2.1 for hospitalization] & 0.01 [OR = 2.7 for ICU]	(Vietzen et al., 2021)
Italian	332 severe patients 1668 controls	32 hemostatic genes ***	*PROC* chr2:127192625:G:A *MTHFR* chr1:11753033:G:A *MTR* chr1:237145686:A:G *ADAMTS13* chr9:133179750:G:C *THBS2* chr6:169195156:A:T	8.77×10^{-5} [2.23 (1.50–3.34)] 1.08×10^{-4} [1.56 (1.25–1.95)] 3.14×10^{-4} [3.49 (1.77–6.88)] 4.26×10^{-4} [2.90 (1.60–5.24)] 7.84×10^{-4} [2.35 (1.43–3.87)]	(Cappadona et al., 2021)
Italian	332 severe patients 1668 controls	*MBL2*	Haplotype composed of rs10824844, and rs10824845	1.04×10^{-5} [1.88 (1.44–2.45)]	(Stravalaci et al., 2022)
Italian	527 severe patients 3190 controls	*MBL2* ****	Haplotype composed of rs17662822, rs1159798, and rs1912619	2.96×10^{-5} [0.11 (0.046–0.27)]	(Asselta et al., 2022)
German	123 cases 94 healthy controls	30 identified as associated from published data	*LZTFL1* rs73064425	0.0004 [7.75 (2.1–28.55)]	(Rüter et al., 2022)
Not reported	284 patients 100 healthy controls	*MBL2*	rs1800450 *****	BB genotype $P = 0.001$ [severity OR = 5.3] & AB genotype $P = 0.001$ [severity OR = 2.9]	(Medetalibeyoglu et al., 2021)
Iranian	375 survivor cases 375 non survivor cases	*IFNL3*, *IFNL4*	*IFNL3* rs12979860 *IFNL3* rs8099917 *IFNL3* rs12980275 *IFNL4* rs368234815	CC genotype <0.001 [0.74 (0.038–0.94)] TT genotype <0.001 [0.29 (0.17–0.49)] AA genotype <0.001 [0.42 (0.26–0.68)] TT/TT genotype <0.001 [0.12 (0.067–0.22)]	(Rahimi et al., 2021)
Iranian	186 severe cases 102 non severe cases	*TMPRSS2*	rs17854725 rs12329760	AA vs. AG & AA vs. GG <0.001 CC vs. TC <0.001	(Rokni et al., 2022)

Only studies with at least 100 patients and with the most significant association at ≤0.01 were included. Genes replicated at least once in candidate-gene association studies are underlined.
* We reported all the available data; ** rs429358 and rs7412 were used to infer the *APOE* e4e4, e3e4, and e3e3 haplotypes; *** only the top 5 hits are reported; **** only the top signal is reported; ***** legacy names for the two alleles are A and B (genotypes are indicated as AA, AB and BB).

risk profile that is strongly associated with the severity of the disease among cases, but modestly improves the prediction of disease severity when compared to demographic/clinical factors alone (Horowitz et al., 2022; Dite et al., 2021). Further models, possibly including a higher number of genetic predisposing factors, are hence necessary to be developed.

Notwithstanding all strategies adopted so far, a big proportion of heritability is still to be disclosed. In this respect, the heritability for COVID-19 symptoms explained by common variants was estimated to be 6.5% (The GenOMICC Investigators et al., 2021), whereas the phenotypic variance due to host genetic factors of predicted COVID-19 is around 31% (as deduced in a twin study with participants from the TwinsUK cohort) (Williams et al., 2020). This means that we still have nearly 25% of unexplained heritability. Considering that the current studies on rare variants, unlike for many complex diseases and traits, do not add a major contribution to COVID-19 severity missing heritability, we are in the urgent need of alternative approaches to unravel this genetic dark side.

Numerous alternative techniques and strategies could help the field to move forward: for instance, epistatic (gene*gene), interaction (gene*environment), or haplotype analyses could be attempted. Not to mention the novel and extremely promising results of artificial-intelligence approaches to human genetics (Gerussi et al., 2022; De La Vega et al., 2021; Yang et al., 2021). Haplotype analysis seems to be an old strategy that still deserves to be explored. In this respect, a paradigmatic example could be represented by the *MBL2* gene: besides strong functional data supporting the role of this gene in the pathogenesis of COVID-19, a genetic study only based on single-variant association analysis would have produced only marginal results. Instead, Stravalaci and colleagues, by performing an extensive haplotype analysis of the *MBL2* locus, were able to evidence genome-wide significant association signals, suggesting that such an approach, although computationally important, could highlight the contribution of genes that would otherwise remain undetected (Stravalaci et al., 2022; Asselta et al., 2022).

Finally, for therapeutic purposes, extensive follow up of both common and rare variants should be considered, especially in the light of the fact that two-thirds of FDA-approved drugs in 2021 have a clear genetic evidence (Ochoa et al., 2022).

Funding

This work was supported by the generous contribution of Banca Intesa San Paolo, and by EU funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT).

CRediT authorship contribution statement

Claudio Cappadona: Data curation, Writing – review & editing, Visualization. **Valeria Rimoldi:** Data curation, Writing – review & editing. **Elvezia Maria Paraboschi:** Data curation, Writing – review & editing. **Rosanna Asselta:** Conceptualization, Data curation, Writing – original draft, Supervision, Funding acquisition.

Declaration of Competing Interest

All Authors of the Review entitled "***Genetic susceptibility to severe COVID-19***", declare NO conflicts of interests.

Data availability

No data was used for the research described in the article.

Acknowledgments

This manuscript is dedicated in memory of Prof. Stefano Duga, whose enormous support and guidance had inspired, and continue to inspire us immensely.

References

Agwa, S.H.A., Kamel, M.M., Elghazaly, H., Abd Elsamee, A.M., Hafez, H., Girgis, S.A., et al., 2021. Association between interferon-Lambda-3 rs12979860, TLL1 rs17047200 and DDR1 rs4618569 variant polymorphisms with the course and outcome of SARS-CoV-2 patients. Genes. 12 (6), 830.

Alghamdi, J., Alaamery, M., Barhoumi, T., Rashid, M., Alajmi, H., Aljasser, N., et al., 2021. Interferon-induced transmembrane protein-3 genetic variant rs12252 is associated with COVID-19 mortality. Genomics 113 (4), 1733–1741.

Asano, T., Boisson, B., Onodi, F., Matuozzo, D., Moncada-Velez, M., Maglorius Renkilaraj, M.R.L., et al., 2021. X-linked recessive TLR7 deficiency in ~1% of men under 60 years old with life-threatening COVID-19. Sci. Immunol. 6 (62). , eabl4348.

Asselta, R., Paraboschi, E.M., Stravalaci, M., Invernizzi, P., Bonfanti, P., Biondi, A., et al., 2022. Reply to: Hultström et al., genetic determinants of mannose-binding lectin activity predispose to thromboembolic complications in critical COVID-19. Mannose-binding lectin genetics in COVID-19. Nat. Immunol. 23 (6), 865–867.

Bastard, P., Rosen, L.B., Zhang, Q., Michailidis, E., Hoffmann, H.H., Zhang, Y., et al., 2020. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. Science. 370 (6515) eabd4585.

Bouayad, A., 2020. Innate immune evasion by SARS-CoV -2: comparison with SARS-CoV. Rev. Med. Virol 30 (6), 1–9.

Butler-Laporte, G., Povysil, G., Kosmicki, J.A., Cirulli, E.T., Drivas, T., Furini, S., et al., 2022. Exome-wide association study to identify rare variants influencing COVID-19 outcomes: Results from the Host Genetics Initiative. Cooper GM, curatore. PLoS Genet. 18 (11), e1010367.

Cappadona, C., Paraboschi, E.M., Ziliotto, N., Bottaro, S., Rimoldi, V., Gerussi, A., et al., 2021. MEDTEC students against coronavirus: investigating the role of hemostatic genes in the predisposition to COVID-19 severity. J. Pers Med. 11 (11), 1166.

Chen, T., Wu, D., Chen, H., Yan, W., Yang, D., Chen, G., et al., 2020. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. BMJ. m1091.

COVID-19 Host Genetics Initiative, 2021. COVID-19 Host Genetics InitiativeLeadership. In: MEK, Niemi, Karjalainen, J., Liao, R.G., Neale, B.M., et al. (Eds.), Mapping the human genetic architecture of COVID-19. Nature, 600(7889). pp. 472–477.

COVID-19 Host Genetics Initiative, 2022. COVID-19 Host Genetics Initiative, Leadership. In: Pathak, G.A., Karjalainen, J., Stevens, C., et al. (Eds.), A First Update on Mapping the Human Genetic Architecture of COVID-19. Nature. 608 (7921), E1–10.

COVID-19 Host Genetics Initiative, COVID-19 Host Genetics InitiativeLeadership, MEK, Niemi, Karjalainen, J., Liao, R.G., Neale, B.M., et al., 2021. Mapping the human genetic architecture of COVID-19. Nature 600 (7889), 472–477.

Cruz, R., Diz-de Almeida, S., López de Heredia, M., Quintela, I., Ceballos, F.C., Pita, G., et al., 2022. Novel genes and sex differences in COVID-19 severity. Hum. Mol. Genet. 31 (22), 3789–3806.

da Silva, S.J.R., do Nascimento, J.C.F., Germano Mendes, R.P., Guarines, K.M., Targino Alves da Silva, C., da Silva, P.G., et al., 2022. Two years into the COVID-19 pandemic: lessons learned. ACS. Infect. Dis. Ther. 8 (9), 1758–1814.

De La Vega, F.M., Chowdhury, S., Moore, B., Frise, E., McCarthy, J., Hernandez, E.J., et al., 2021. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. Genome Med. 13 (1), 153.

Degenhardt, F., Ellinghaus, D., Juzenas, S., Lerga-Jaso, J., Wendorff, M., Maya-Miles, D., et al., 2022. Detailed stratified GWAS analysis for severe COVID-19 in four European populations. Hum. Mol. Genet. 31 (23), 3945–3966.

Dieter, C., de Brondani, L.A., Leitão, C.B., Gerchman, F., Lemos, N.E., Crispim, D., 2022. Genetic polymorphisms associated with susceptibility to COVID-19 disease and severity: A systematic review and meta-analysis. Novelli G, curatore. PLoS One 17 (7) e0270627.

Dite, G.S., Murphy, N.M., Allman, R., 2021. An integrated clinical and genetic model for predicting risk of severe COVID-19: A population-based case–control study. Novelli G, curatore. PLoS One 16 (2), e0247205.

Downes, D.J., Cross, A.R., Hua, P., Roberts, N., Schwessinger, R., Cutler, A.J., et al., 2021. Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. Nat Genet 53 (11), 1606–1615.

Driggin, E., Madhavan, M.V., Bikdeli, B., Chuich, T., Laracy, J., Biondi-Zoccai, G., et al., 2020. Cardiovascular considerations for patients, health care workers, and health systems during the COVID-19 pandemic. J. Am Coll. Cardiol 75 (18), 2352–2371.

The COVID-19 Host Genetics Initiative, 2020. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur. J. Hum. Genet. 28 (6), 715–718.

Fajgenbaum, D.C., June, C.H., 2020. Cytokine Storm. Longo DL, curatore. N. Engl. J. Med. 383 (23), 2255–2273.

Fallerini, C., Daga, S., Mantovani, S., Benetti, E., Picchiotti, N., Francisci, D., et al., 2021. Association of Toll-like receptor 7 variants with life-threatening COVID-19 disease in males: findings from a nested case-control study. eLife 10 e67569.

Fitipaldi, H., Franks, P.W., 2022. Ethnic, gender and other sociodemographic biases in genome-wide association studies for the most burdensome non-communicable diseases: 2005–2022. Hum. Mol. Genet, ddac245.

Gerussi, A., Verda, D., Cappadona, C., Cristoferi, L., Bernasconi, D.P., Bottaro, S., et al., 2022. LLM-PBC: logic learning machine-based explainable rules accurately stratify the genetic risk of primary biliary cholangitis. J. Pers Med. 12 (10), 1587.
Gong, B., Huang, L., He, Y., Xie, W., Yin, Y., Shi, Y., et al., 2022. A genetic variant in IL-6 lowering its expression is protective for critical patients with COVID-19. Signal Transduct. Target Ther. 7 (1), 112.
Grasselli, G., Zangrillo, A., Zanella, A., Antonelli, M., Cabrini, L., Castelli, A., et al., 2020. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy region, Italy. JAMA. 323 (16), 1574.
Hariharan, A., Hakeem, A.R., Radhakrishnan, S., Reddy, M.S., Rela, M., 2021. The role and therapeutic potential of NF-kappa-B pathway in severe COVID-19 patients. Inflammopharmacology 29 (1), 91–100.
Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al., 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181 (2), 271–280.e8.
Horowitz, J.E., Kosmicki, J.A., Damask, A., Sharma, D., Roberts, G.H.L., Justice, A.E., et al., 2022. Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. Nat. Genet 54 (4), 382–392.
Ishak, A., Mehendale, M., AlRawashdeh, M.M., Sestacovschi, C., Sharath, M., Pandav, K., et al., 2022. The association of COVID-19 severity and susceptibility and genetic risk factors: a systematic review of the literature. Gene. 836, 146674.
Kasuga, Y., Zhu, B., Jang, K.J., Yoo, J.S., 2021. Innate immune sensing of coronavirus and viral evasion strategies. Exp Mol Med 53 (5), 723–736.
Knoops, K., Kikkert, M., van den Worm, S.H.E., Zevenhoven-Dobbe, J.C., van der Meer, Y., Koster, A.J., et al., 2008. SARS-Coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. Emerman M, curatore. PLoS Biol. 6 (9), e226.
Kosmicki, J.A., Horowitz, J.E., Banerjee, N., Lanche, R., Marcketta, A., Maxwell, E., et al., 2021. Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586,157 individuals. Am. J. Hum. Genet. 108 (7), 1350–1355.
Kousathanas, A., Pairo-Castineira, E., Rawlik, K., Stuckey, A., Odhams, C.A., Walker, S., et al., 2022. Whole-genome sequencing reveals host factors underlying critical COVID-19. Nature. 607 (7917), 97–103.
Kuo, C.L., Pilling, L.C., Atkins, J.L., Masoli, J.A.H., Delgado, J., Kuchel, G.A., et al., 2020. APOE e4 genotype predicts severe COVID-19 in the UK Biobank community cohort. J. Gerontol. Ser. A. 75 (11), 2231–2232.
Li, Y., Ke, Y., Xia, X., Wang, Y., Cheng, F., Liu, X., et al., 2021. Genome-wide association study of COVID-19 severity among the Chinese population. Cell Discov. 7 (1), 76.
López-Pacheco, C., Soldevila, G., Du Pont, G., Hernández-Pando, R., García-Zepeda, E.A., 2016. CCR9 is a key regulator of early phases of allergic airway inflammation. Mediat. Inflamm 1–16.
Medetalibeyoglu, A., Bahat, G., Senkal, N., Kose, M., Avci, K., Sayin, G.Y., et al., 2021. Mannose binding lectin gene 2 (rs1800450) missense variant may contribute to development and severity of COVID-19 infection. Infect. Genet. Evol. 89, 104717.
Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., Manson, J.J., 2020. COVID-19: consider cytokine storm syndromes and immunosuppression. The Lancet 395 (10229), 1033–1034.
Merad, M., Martin, J.C., 2020. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat. Rev. Immunol. 20 (6), 355–362.
Möhlendick, B., Schönfelder, K., Breuckmann, K., Elsner, C., Babel, N., Balfanz, P., et al., 2021. ACE2 polymorphism and susceptibility for SARS-CoV-2 infection and severity of COVID-19. Pharmacogenet Genom 31 (8), 165–171.
Moreno-Pérez, O., Merino, E., Leon-Ramirez, J.M., Andres, M., Ramos, J.M., Arenas-Jiménez, J., et al., 2021. Post-acute COVID-19 syndrome. Incidence and risk factors: a Mediterranean cohort study. J. Infect 82 (3), 378–383.
The Severe Covid-19 GWAS Group, 2020. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N. Engl. J. Med. 383 (16), 1522–1534.
Nakanishi, T., Pigazzini, S., Degenhardt, F., Cordioli, M., Butler-Laporte, G., Maya-Miles, D., et al., 2021. Age-dependent impact of the major common genetic risk factor for COVID-19 on severity and mortality. J. Clin. Invest. 131 (23), e152386.
Namkoong, H., Edahiro, R., Takano, T., Nishihara, H., Shirai, Y., Sonehara, K., et al., 2022. DOCK2 is involved in the host genetics and biology of severe COVID-19. Nature. 609 (7928), 754–760.
Niemi, M.E.K., Daly, M.J., Ganna, A., 2022. The human genetic epidemiology of COVID-19. Nat. Rev. Genet. 23 (9), 533–546.
Ochoa, D., Karim, M., Ghoussaini, M., Hulcoop, D.G., McDonagh, E.M., Dunham, I., 2022. Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. Nat. Rev. Drug Discov. 21 (8), 551.
Oscanoa, J., Sivapalan, L., Gadaleta, E., Dayem Ullah, A.Z., Lemoine, N.R., Chelala, C., 2020. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic Acids Res. 48 (W1), W185–W192.
Osuchowski, M.F., Winkler, M.S., Skirecki, T., Cajander, S., Shankar-Hari, M., Lachmann, G., et al., 2021. The COVID-19 puzzle: deciphering pathophysiology and phenotypes of a new disease entity. Lancet Respir Med 9 (6), 622–642.
Pandit, R., Singh, I., Ansari, A., Raval, J., Patel, Z., Dixit, R., et al., 2022. First report on genome wide association study in western Indian population reveals host genetic factors for COVID-19 severity and outcome. Genomics 114 (4), 110399.
Pereira, A.C., Bes, T.M., Velho, M., Marques, E., Jannes, C.E., Valino, K.R., et al., 2022. Genetic risk factors and COVID-19 severity in Brazil: results from BRACOVID study. Hum. Mol. Genet. 31 (18), 3021–3031.
Petrilli, C.M., Jones, S.A., Yang, J., Rajagopalan, H., O'Donnell, L., Chernyak, Y., et al., 2020. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in new York City: prospective cohort study. BMJ, m1966.
Povysil, G., Butler-Laporte, G., Shang, N., Wang, C., Khan, A., Alaamery, M., et al., 2021. Rare loss-of-function variants in type I IFN immunity genes are not associated with severe COVID-19. J. Clin. Invest. 131 (14). , e147834.
Rahimi, P., Tarharoudi, R., Rahimpour, A., Mosayebi Amroabadi, J., Ahmadi, I., Anvari, E., et al., 2021. The association between interferon lambda 3 and 4 gene single-nucleotide polymorphisms and the recovery of COVID-19 patients. Virol J. 18 (1), 221.
Rendeiro, A.F., Ravichandran, H., Bram, Y., Chandar, V., Kim, J., Meydan, C., et al., 2021. The spatial landscape of lung pathology during COVID-19 progression. Nature. 593 (7860), 564–567.
Roberts, G.H.L., Partha, R., Rhead, B., Knight, S.C., Park, D.S., Coignet, M.V., et al., 2022. Expanded COVID-19 phenotype definitions reveal distinct patterns of genetic association and protective effects. Nat. Genet 54 (4), 374–381.
Rokni, M., Heidari Nia, M., Sarhadi, M., Mirinejad, S., Sargazi, S., Moudi, M., et al., 2022. Association of TMPRSS2 gene polymorphisms with COVID-19 severity and mortality: a case-control study with computational analyses. Appl. Biochem. Biotechnol. 194 (8), 3507–3526.
Rüter, J., Pallerla, S.R., Meyer, C.G., Casadei, N., Sonnabend, M., Peter, S., et al., 2022. Host genetic loci LZTFL1 and CCL2 associated with SARS-CoV-2 infection and severity of COVID-19. Int. J. Infect. Dis. 122, 427–436.
Schneider, W.M., Chevillotte, M.D., Rice, C.M., 2014. Interferon-stimulated genes: a complex web of host defenses. Annu. Rev. Immunol. 32 (1), 513–545.
Schultze, J.L., Aschenbrenner, A.C., 2021. COVID-19 and the human innate immune system. Cell 184 (7), 1671–1692.
Shelton, J.F., Shastri, A.J., Ye, C., Weldon, C.H., Filshtein-Sonmez, T., Coker, D., et al., 2021. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. Nat Genet 53 (6), 801–808.
Stravalaci, M., Pagani, I., Paraboschi, E.M., Pedotti, M., Doni, A., Scavello, F., et al., 2022. Recognition and inhibition of SARS-CoV-2 by humoral innate immunity pattern recognition molecules. Nat. Immunol. 23 (2), 275–286.
Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., et al., 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 49 (D1), D605–D612.
Tan, L.Y., Komarasamy, T.V., RMT Balasubramaniam V., 2021. Hyperinflammatory immune response and COVID-19: a double edged sword. Front. Immunol. 12, 742941.
Terpos, E., Ntanasis-Stathopoulos, I., Elalamy, I., Kastritis, E., Sergentanis, T.N., Politou, M., et al., 2020. Hematological findings and complications of COVID -19. Am. J. Hematol 95 (7), 834–847.
The GenOMICC Investigators, The ISARIC4C Investigators, The COVID-19 Human Genetics Initiative, 23andMe Investigators, BRACOVID Investigators, Gen-COVID Investigators, et al., 2021. Genetic mechanisms of critical illness in COVID-19. Nature. 591 (7848), 92–98.
Thorne, L.G., Reuschl, A.K., Zuliani-Alvarez, L., Whelan, M.V.X., Turner, J., Noursadeghi, M., et al., 2021. SARS-CoV-2 sensing by RIG-I and MDA5 links epithelial infection to macrophage inflammation. EMBO J 40 (15), e107826.
Tsang, K.W., Ho, P.L., Ooi, G.C., Yee, W.K., Wang, T., Chan-Yeung, M., et al., 2003. A cluster of cases of severe acute respiratory syndrome in Hong Kong. N. Engl. J. Med. 348 (20), 1977–1985.
van der Made, C.I., Simons, A., Schuurs-Hoeijmakers, J., van den Heuvel, G., Mantere, T., Kersten, S., et al., 2020. Presence of genetic variants among young men with severe COVID-19. JAMA. 324 (7), 663–673.
Verma, S., Abbas, M., Verma, S., Khan, F.H., Raza, S.T., Siddiqi, Z., et al., 2021. Impact of I/D polymorphism of angiotensin-converting enzyme 1 (ACE1) gene on the severity of COVID-19 patients. Infect. Genet. Evol 91,, 104801.
Vietzen, H., Zoufaly, A., Traugott, M., Aberle, J., Aberle, S.W., Puchhammer-Stöckl, E., 2021. Deletion of the NKG2C receptor encoding KLRC2 gene and HLA-E variants are risk factors for severe COVID-19. Genet Med 23 (5), 963–967.
Vuille-dit-Bille, R.N., Camargo, S.M., Emmenegger, L., Sasse, T., Kummer, E., Jando, J., et al., 2015. Human intestine luminal ACE2 and amino acid transporter expression increased by ACE-inhibitors. Amino Acids 47 (4), 693–705.
Wein, A.N., McMaster, S.R., Takamura, S., Dunbar, P.R., Cartwright, E.K., Hayward, S.L., et al., 2019. CXCR6 regulates localization of tissue-resident memory CD8 T cells to the airways. J. Exp. Med. 216 (12), 2748–2762.
Williams, F.M.K., Freidin, M.B., Mangino, M., Couvreur, S., Visconti, A., Bowyer, R.C.E., et al., 2020. Self-reported symptoms of COVID-19, including symptoms Most predictive of SARS-CoV-2 infection, are heritable. Twin Res Hum Genet. 23 (6), 316–321.
Wu, Z., McGoogan, J.M., 2020. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA. 323 (13), 1239.
Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., et al., 2020. A new coronavirus associated with human respiratory disease in China. Nature. 579 (7798), 265–269.
Wu, P., Ding, L., Li, X., Liu, S., Cheng, F., He, Q., et al., 2021. Trans-ethnic genome-wide association study of severe COVID-19. Commun Biol. 4 (1), 1034.
Yang, N.I., Yeh, C.H., Tsai, T.H., Chou, Y.J., Hsu, P.W.C., Li, C.H., et al., 2021. Artificial intelligence-assisted identification of genetic factors predisposing high-risk individuals to asymptomatic heart failure. Cells. 10 (9), 2430.
Zeberg, H., Pääbo, S., 2020. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. Nature. 587 (7835), 610–612.

Zecevic, M., Kotur, N., Ristivojevic, B., Gasic, V., Skodric-Trifunovic, V., Stjepanovic, M., et al., 2022. Genome-wide association study of COVID-19 outcomes reveals novel host genetic risk loci in the Serbian population. Front. Genet. 13, 911010.

Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., et al., 2020. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science. 370 (6515) eabd4570.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al., 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382 (8), 727–733.